



National Institutes
of Health

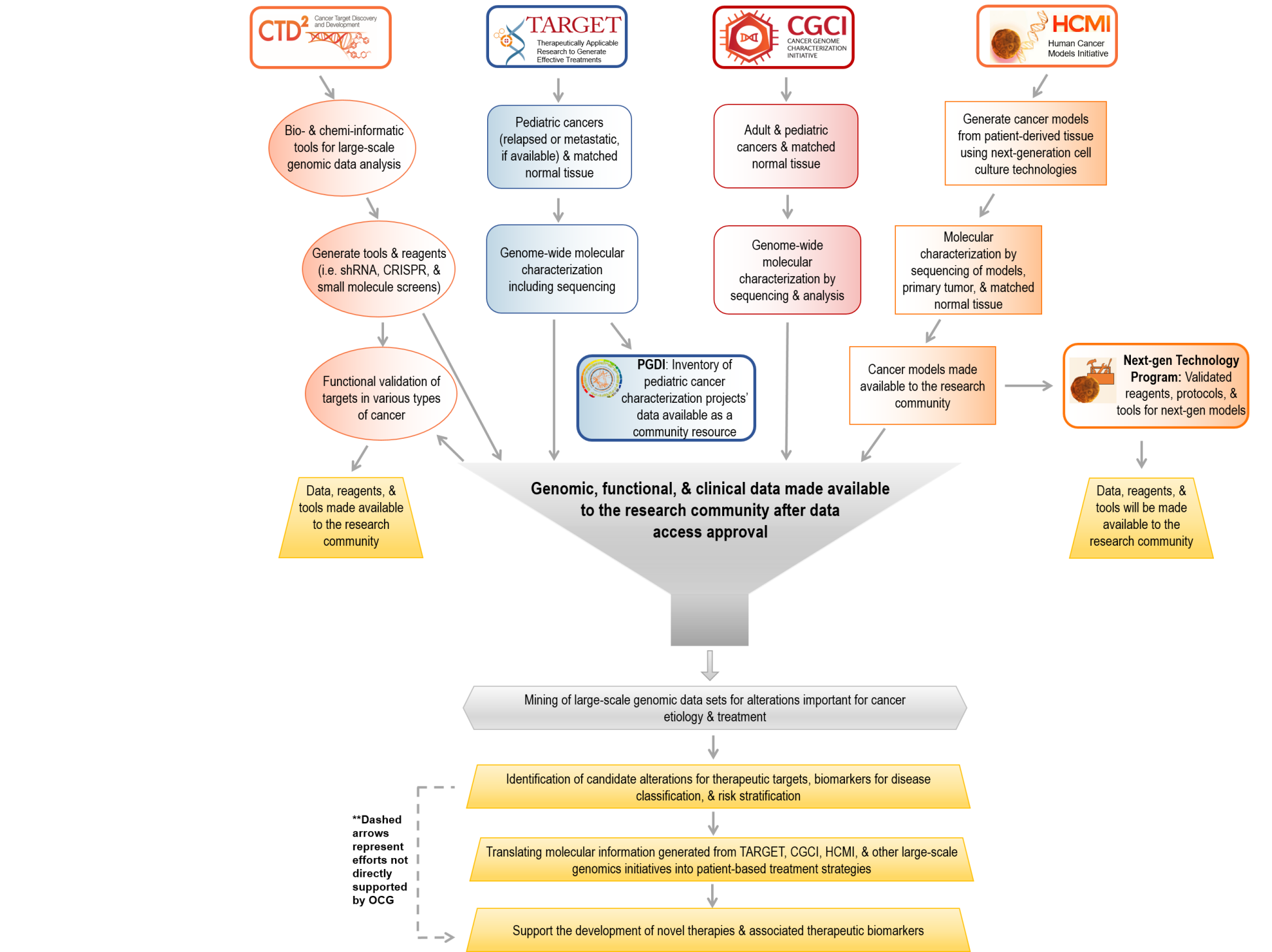
NCI OFFICE OF CANCER GENOMICS: MULTIDISCIPLINARY GENOMIC RESEARCH INITIATIVES DRIVE DISCOVERIES TOWARDS THE CLINIC

Cindy W. Kyi, Pamela C. Birriel, Tanja M. Davidsen, Martin L. Ferguson, Patee Gesuwan, Nicholas B. Griner, Yiwen He, Julyann Perez-Mayoral, Eva Tonsing-Carter, and Daniela S. Gerhard



Our Mission

The National Cancer Institute's [Office of Cancer Genomics](#) (OCG) aims to advance the molecular understanding of cancers and their response to therapies to improve clinical outcomes. To accomplish this goal, OCG develops and collaboratively manages molecular characterization and translational genomic research initiatives. All [OCG programs](#) share data and resources with the research community (as shown in the chart below).



CANCER TARGET DISCOVERY AND DEVELOPMENT (CTD²) NETWORK

The [CTD² Network](#) bridges the gap between genomics and development of effective therapeutics. The Network aims to understand tumor development, heterogeneity, drug resistance, and metastasis to develop therapies combining chemotherapy with immunotherapy.

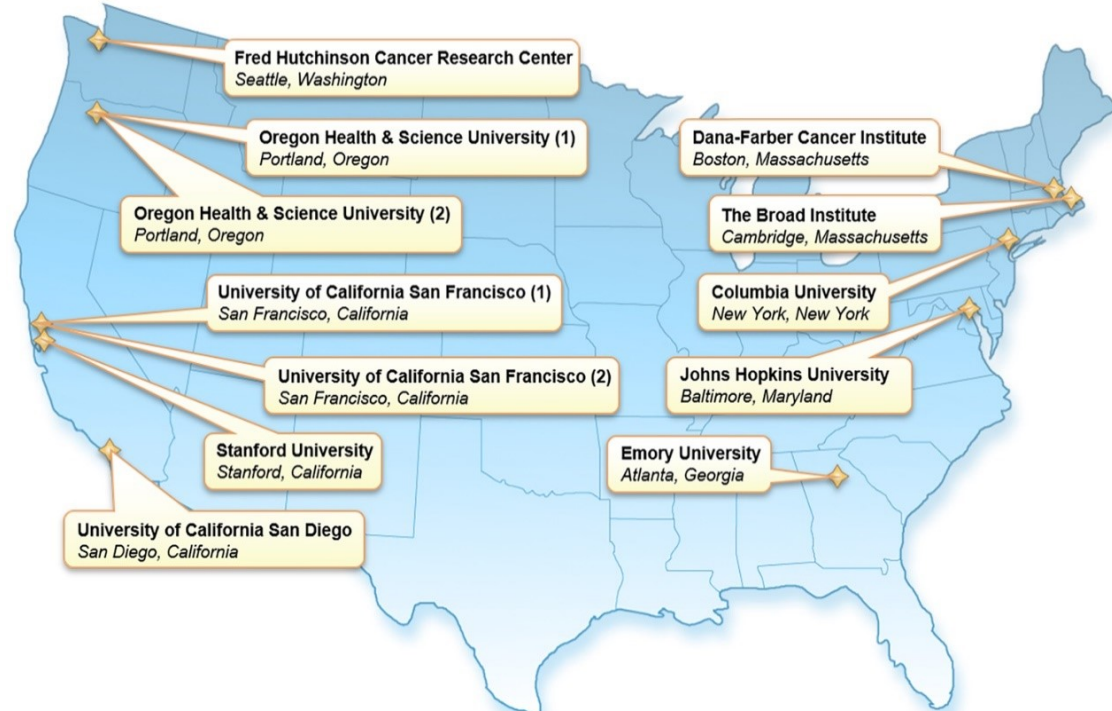


Figure: Map showing CTD² Network Centers.

The Network Centers utilize a distinct array of advanced computational and systems biology methods, functional genomics and immunological approaches, and small molecule and genetic screens. These methods allow reconstruction of cell-context specific gene networks that underlie each cancer subtype, discovery of efficient treatment strategies, and development of optimal combinations of chemical or biological perturbagens.

CANCER GENOME CHARACTERIZATION INITIATIVE (CGCI)

[CGCI](#)-supported research [projects](#) use molecular characterization data to uncover distinct features of rare adult and pediatric cancers. The adult [Burkitt Lymphoma Genome Sequencing Project](#) and [HIV+ Tumor Molecular Characterization Project](#) are two ongoing projects using epigenomic and next-generation sequencing methods.

BURKITT LYMPHOMA GENOME SEQUENCING PROJECT (BLGSP)

The goal of the BLGSP is to explore genomic changes in endemic, sporadic (pediatric and adult), and HIV-associated Burkitt Lymphoma (BL). The results from pediatric BL cases were published in [“Genome-wide discovery of somatic coding and non-coding mutations in pediatric endemic and sporadic Burkitt lymphoma” \(PMID: 30617194\)](#).

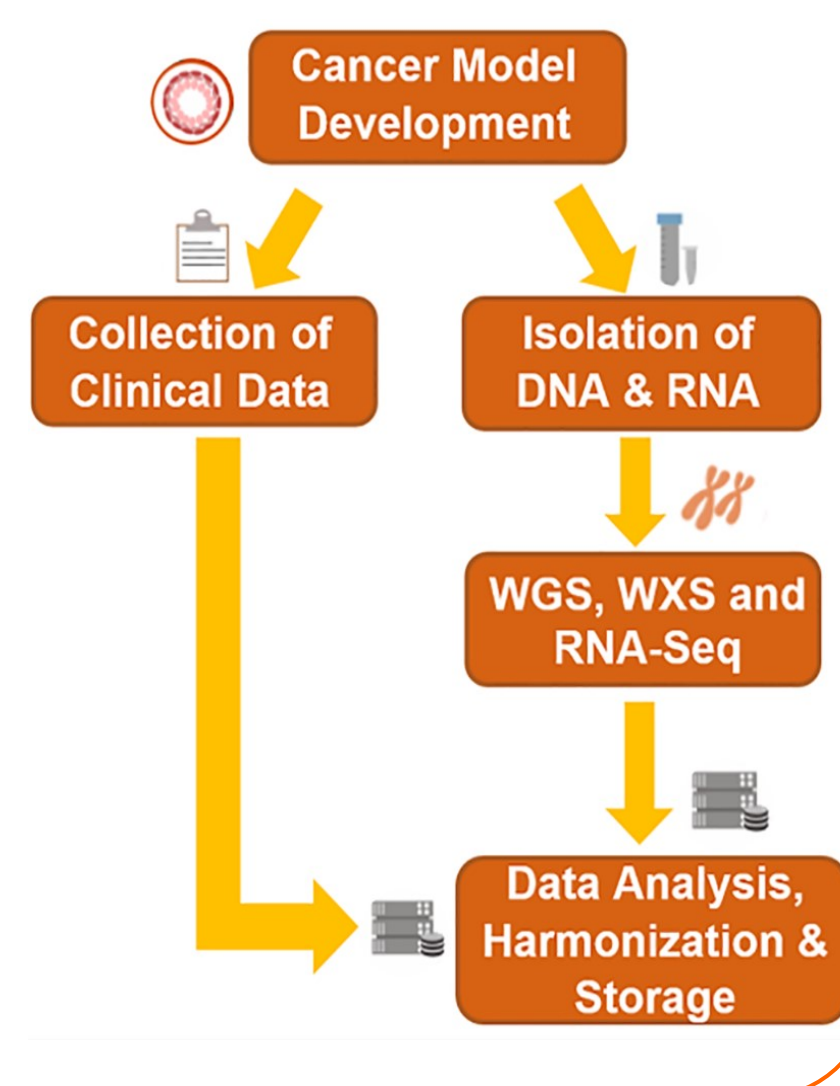
HIV+ TUMOR MOLECULAR CHARACTERIZATION PROJECT (HTMCP)

HTMCP studies HIV-associated cancers including cervical cancer, lung carcinomas and diffuse large B-cell lymphoma (DLBCL) to uncover distinct genomic features between HIV+ and HIV- patients. HTMCP study [“Analysis of Ugandan cervical carcinomas identifies human papillomavirus clade-specific epigenome and transcriptome landscapes” \(PMID: 32747824\)](#) showed that DNA methylation and epigenetic modification patterns of the tumors differ based on infection of specific human papillomavirus clades.

HUMAN CANCER MODELS INITIATIVE (HCM)

[HCM](#) is an international consortium founded by the National Cancer Institute (NCI), Cancer Research UK, the foundation Hubrecht Organoid Technology, and the Wellcome Sanger Institute. HCM is generating novel, patient-derived next-generation cancer models annotated with clinical, biospecimen and molecular characterization data. The QC'ed models, along with their associated data, are available to researchers as a community resource at NCI's [Genomic Data Commons](#). The models are distributed by a third-party distributor, [American Type Culture Collection](#).

Figure: NCI Cancer Model Development Pipeline. Models are developed by the [Cancer Model Development Centers](#). Model-associated clinical data are submitted to the Clinical Data Center. Once the clinical data is quality controlled (QC'd), the model and normal cells (and tumor tissue if available) are sent for nucleic acid isolation and QC. The DNA and RNA are sequenced, i.e. whole exome (WXS), whole genome (WGS), and RNA (RNA-Seq). The models are also characterized for epigenetic state. The sequencing data are harmonized and analyzed by NCI's Genomic Data Commons (GDC). Clinical, biospecimen, and all sequencing data are stored and available at the GDC.



THERAPEUTICALLY APPLICABLE RESEARCH TO GENERATE EFFECTIVE TREATMENTS (TARGET)

The [TARGET](#) program applies comprehensive genomics approaches to determine molecular changes that drive childhood cancers. The goal of the program is to identify effective therapeutic targets and biomarkers. TARGET is organized into a collaborative network of disease-specific project teams. TARGET investigators analyzed the genome, transcriptome, and epigenome of each cancer subtype to guide development of effective, less toxic therapies that will improve patient outcomes ([see publications](#)).

NEXT-GENERATION TECHNOLOGIES FOR NEXT-GENERATION CANCER MODELS (NGT)

The goal of the [NGT](#) Program ([RFA-CA-19-055](#)) is to support the development of technologies and tools that will accelerate and enhance research using advanced human-derived next-generation cancer models. The tools developed will provide researchers robust and reproducible genome editing/manipulating protocols and reagents and enable advanced data interpretation. Protocols, knowledge, and materials developed will be shared broadly with the research community.

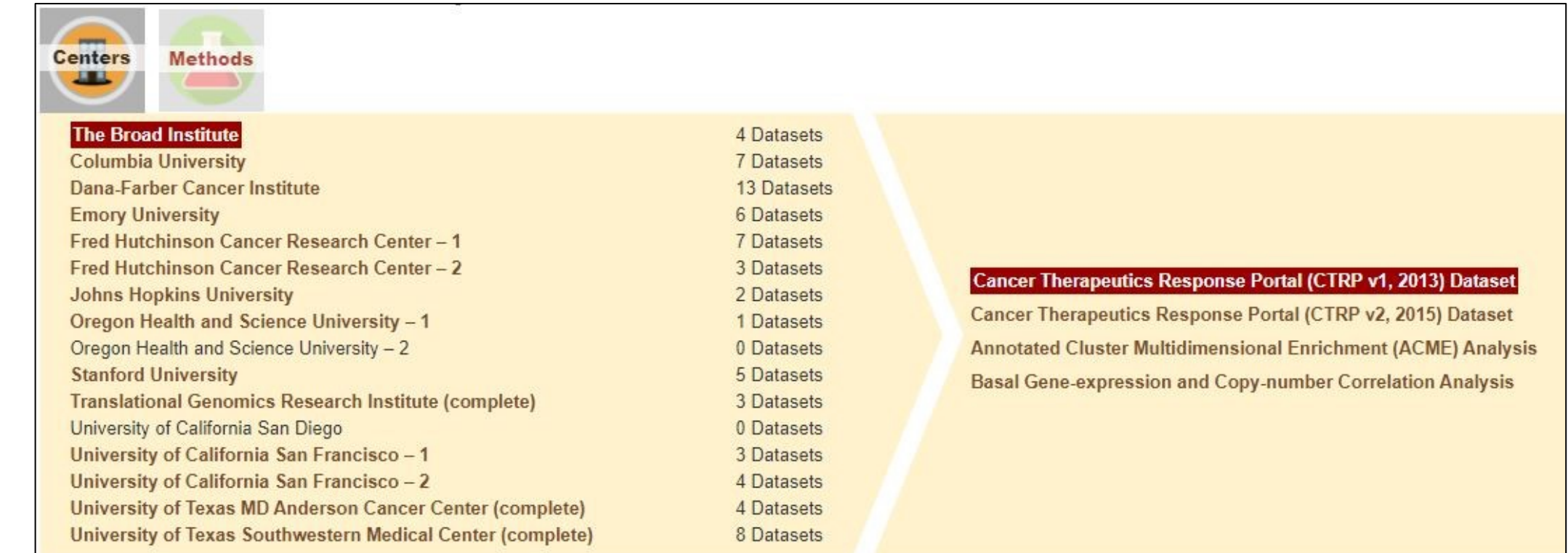
Broad Institute, Dana Farber Cancer Institute, and Massachusetts Institute of Technology are the Centers working on developing the next-generation technologies and tools under this program.

Databases and Resources for the Research Community

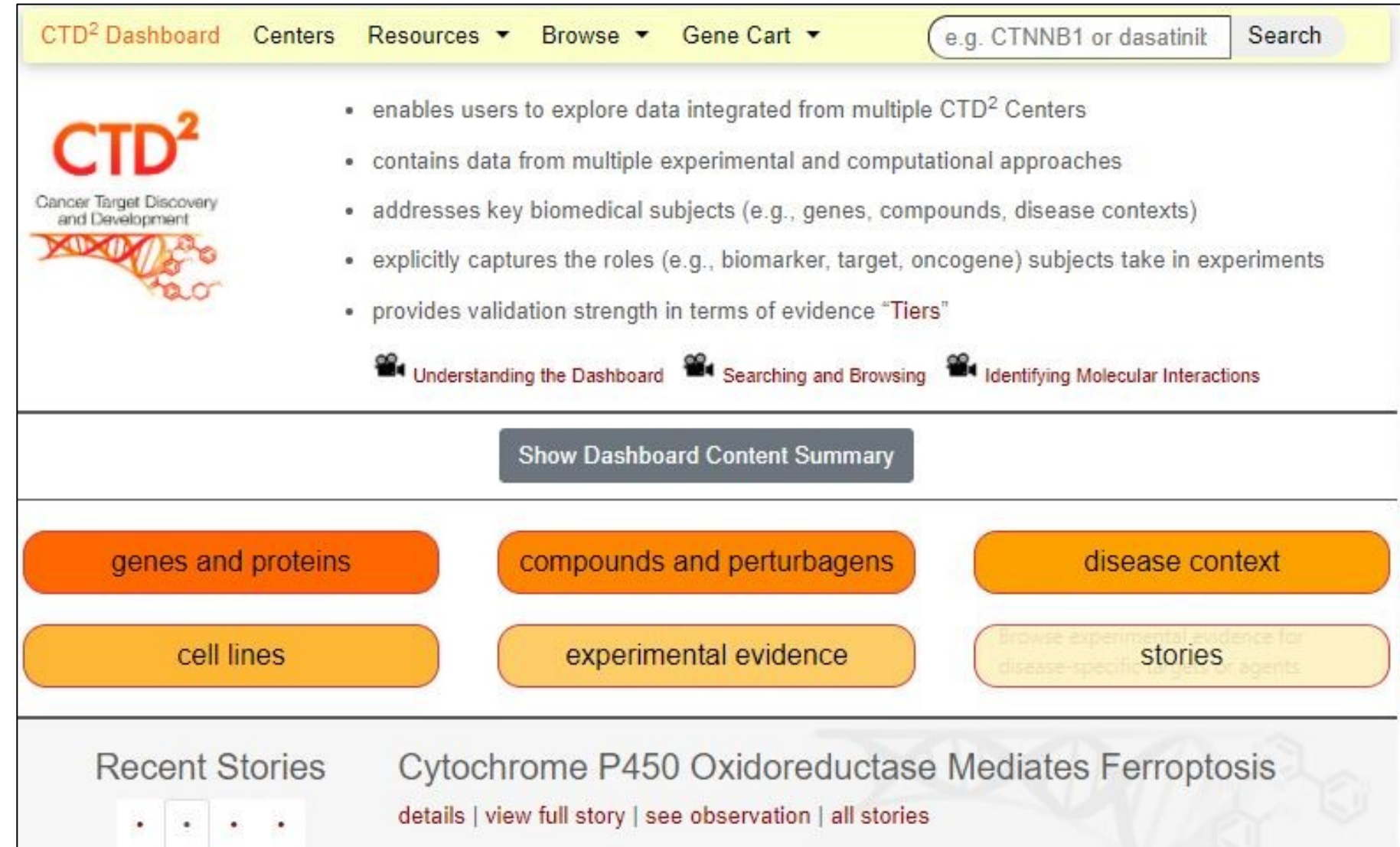
ACCESSING CTD² DATA

Raw and analyzed primary data are available through the [CTD² Data Portal](#). Network Center-generated conclusions or “observations” with associated supporting evidence can be accessed through the interactive web interface, [CTD² Dashboard](#). The Dashboard allows easy navigation to computational and non-computational scientists. All data generated by this Network are open access.

CTD² DATA PORTAL



CTD² DASHBOARD



CTD² ANALYTICAL TOOLS: Access tools that are useful for analyzing and/or visualizing large-scale genomic and/or complex datasets at: <https://ocg.cancer.gov/programs/ctd2/analytical-tools>

CTD² REAGENTS: Materials (e.g. plasmids, cell lines, cell models, etc.) available for use by researchers are listed at: <https://ocg.cancer.gov/programs/ctd2/supported-reagents>

ACCESSING CGCI AND TARGET DATA

OCG's molecular characterization initiatives make their datasets available through user-friendly tabular data matrices at <https://ocg.cancer.gov/programs/target/data-matrix> and <https://ocg.cancer.gov/programs/cgci/data-matrix>. These data matrices provide links to currently available data and metadata stored at OCG's Data Coordinating Center (DCC), National Center for Biotechnology Information (NCBI), and Genomic Data Commons (GDC). New data are added as they are generated and analyzed.

CGCI DATA MATRIX

Disease	Patient Data	Gene Expression	Copy Number	Methylation	mRNA	Sequencing	Other
Burkitt's Lymphoma (BLGSP)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
HIV+ Tumor Molecular Characterization Project (HTMCP)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
HIV+ Cervical Cancers (HTMCP-C)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
HIV+ Lung Cancers (HTMCP-L)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other

Data Legend

Brown: Public*
Red: Protected†
Black: Unavailable

TARGET DATA MATRIX

Disease	Patient Data	Gene Expression	Copy Number	Methylation	mRNA	Sequencing	Other
Acute lymphoblastic leukemia (ALL)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
ALL Phase 1 (B-ALL)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
ALL Phase 2 (B-ALL)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
ALL Phase 3 (B-ALL)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
Acute myeloid leukemia (AML)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
AML	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
Relapsed Leukemia (RL)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
Relapsed Leukemia (RL)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
Child Cell Sarcoma of the Extremity (CCSE)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other
Child Cell Sarcoma of the Extremity (CCSE)	Clinical File Sample Matrix Pathology Images Biopsy Specimen Data	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	mRNA-seq FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Whole Genome FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Targeted Resequencing FASTQ/BAM BCC Open* BCC Controlled† BCC Controlled‡	Other

ACCESSING HCM

HCM Searchable Catalog (<https://hcmi-searchable-catalog.nci.nih.gov/>) is an ongoing resource that allows users to browse and identify available next-generation cancer models. Currently, there are 148 models from 18 primary sites. Visit [HCM Resources](#) for the Catalog user guide.

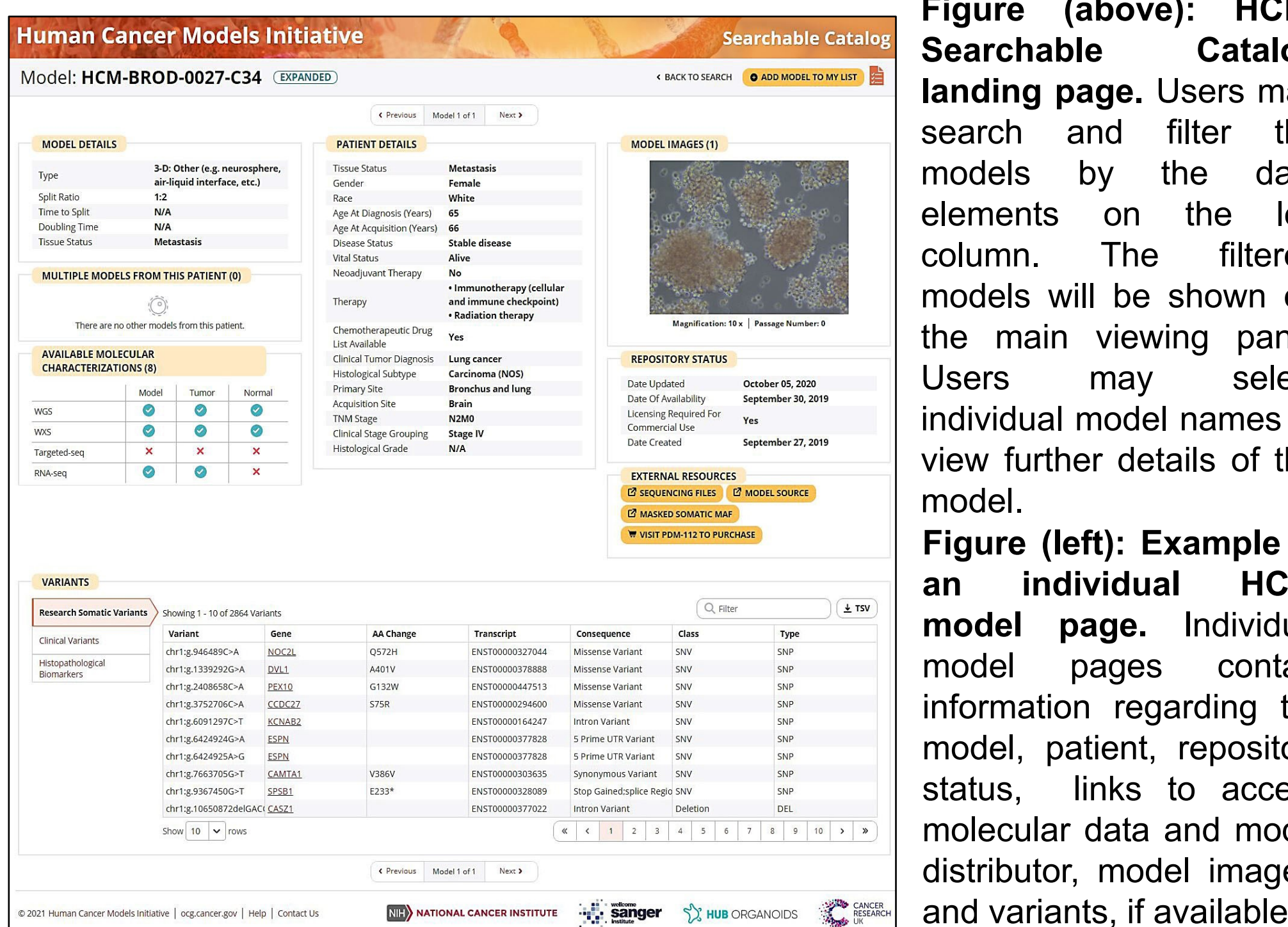
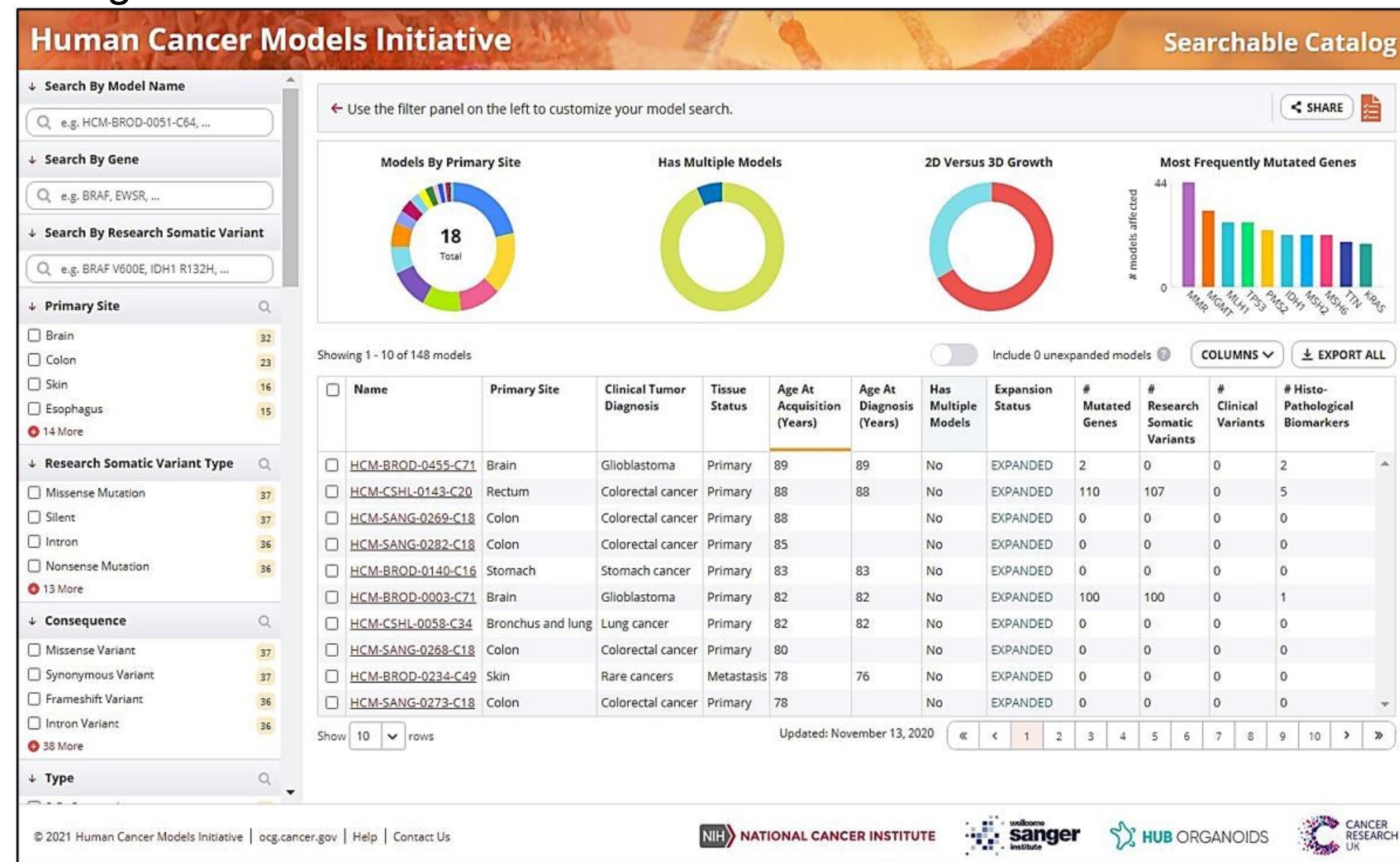


Figure (above): HCM Searchable Catalog landing page. Users may search and filter the models by the data elements on the left column. The filtered models will be shown on the main viewing pane. Users may select individual model names to view further details of the model. **Figure (left): Example of an individual HCM model page.** Individual model pages contain information regarding the model, patient, repository status, links to access molecular data and model distributor, model images, and variants, if available.

HCM CASE REPORT FORMS (CRFs): Cancer type-specific CRFs have been developed through collaborations with international clinical experts. Enrollment and Follow-up CRFs are available for download at: <https://ocg.cancer.gov/programs/hcmi/resources>.

OCG DATA AT GENOMIC DATA COMMONS (GDC)

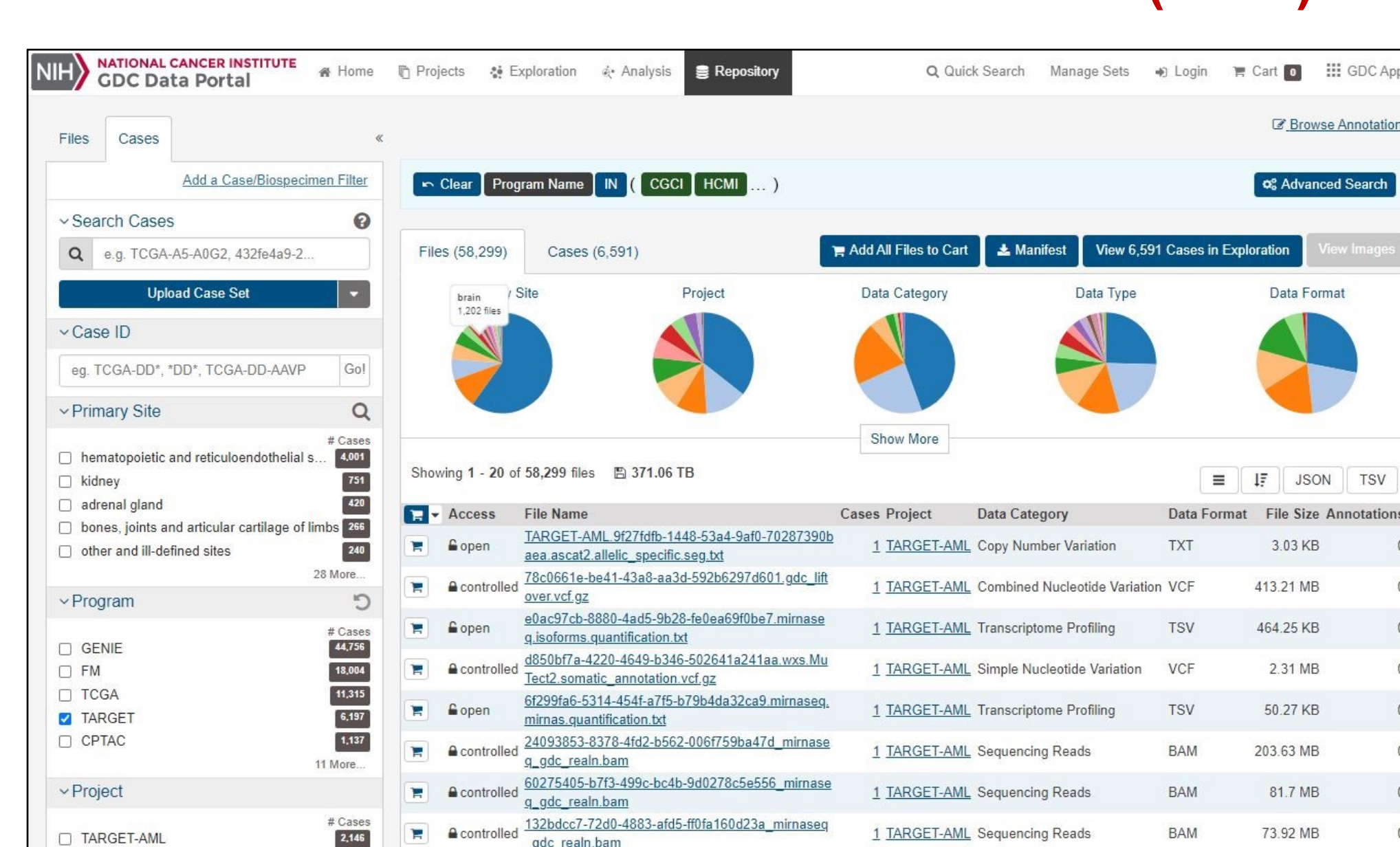


Figure: Snapshot of GDC Data Portal. Users may select available datasets from OCG programs at the GDC by program, disease project, data type, etc. The image displays available open and controlled datasets from TARGET-AML project as an example.

OCG DATA AT DATA COORDINATING CENTER (DCC)

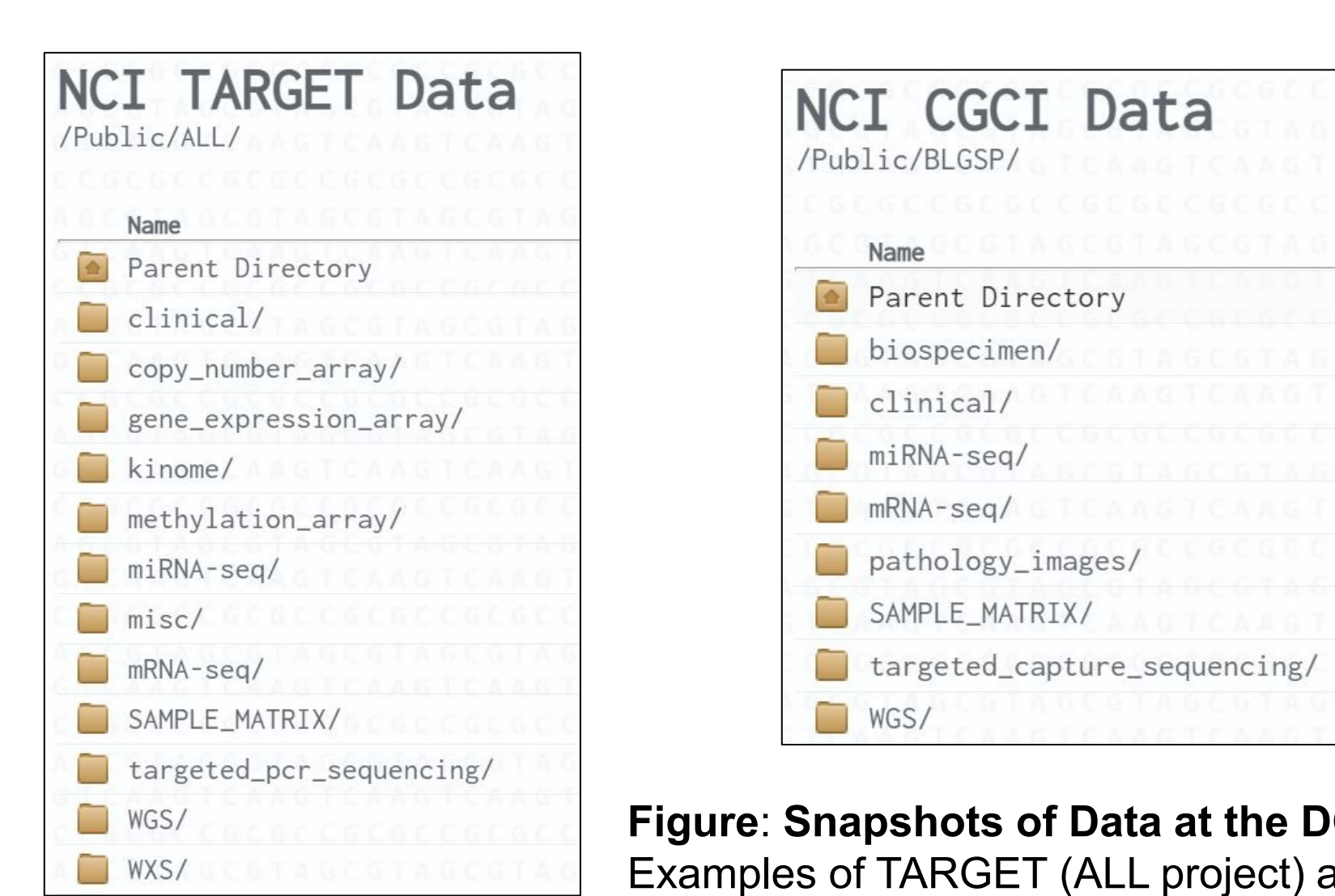


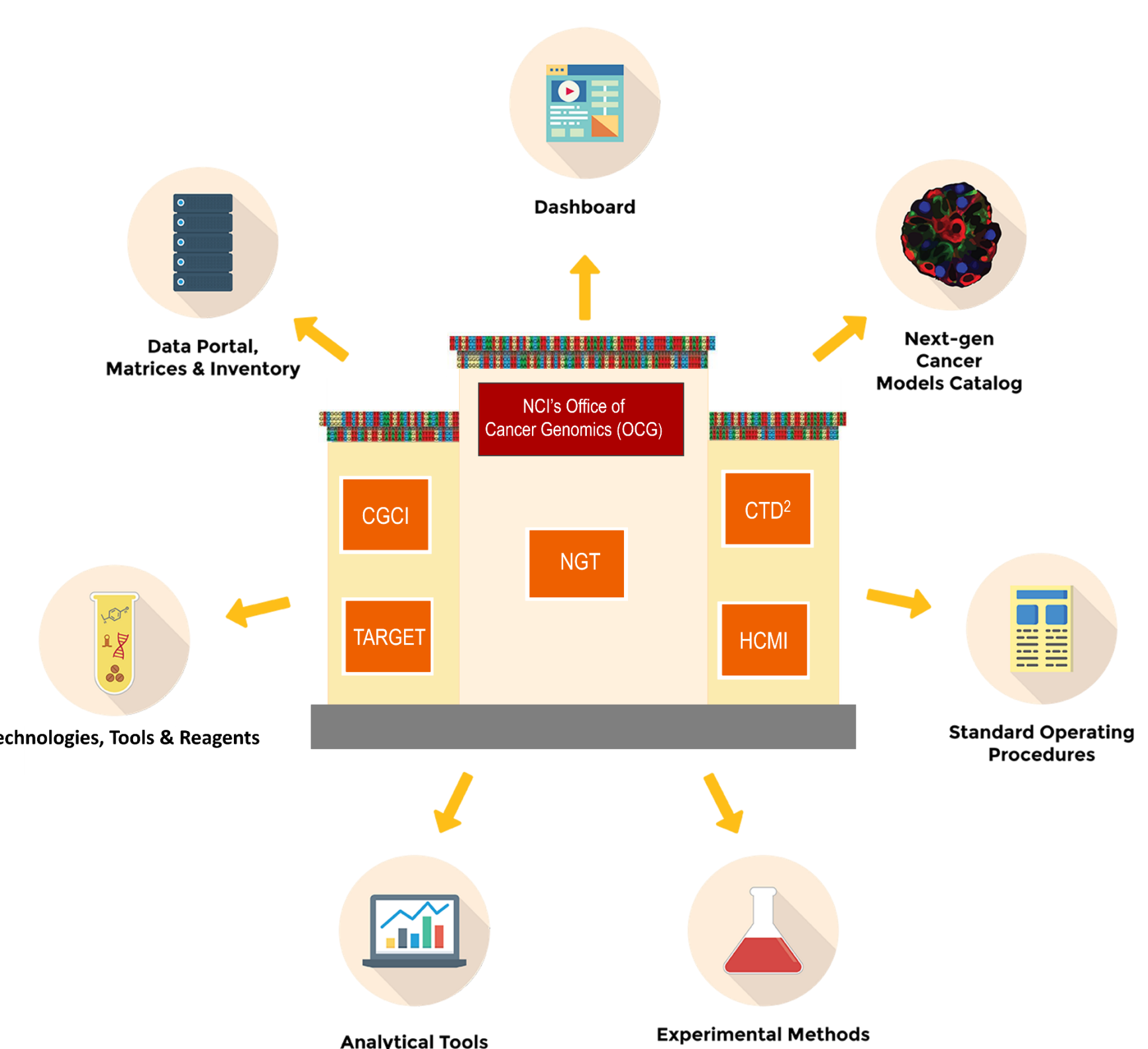
Figure: Snapshots of Data at the DCC. Examples of TARGET (ALL project) and CGCI (BLGSP project) open-access data and metadata available at OCG's DCC.

Visit: ocg.cancer.gov
Sign up for OCG email updates!
OCG@mail.nih.gov

RESOURCES FROM OCG'S PROGRAMS

OCG initiatives promote:

- Generation and dissemination of up-to-date data via programmatic databases such as OCG's DCC, NCBI, and GDC
- Advances in bio- and chemi-informatic methodologies
- Development and sharing of next-generation cancer models, experimental reagents, technologies, analytical tools and standard operating procedures



- OCG program data associated with demographic, clinical, and/or genotypic information may require authorization to access.
- Please read our [data policies](#) and [guide to accessing data](#):
 - <https://ocg.cancer.gov/data/data-policies>
 - <https://ocg.cancer.gov/flowchart/guide-accessing-data>
- We request an acknowledgement of the respective OCG program for data usage according to the guidelines on “Using Data” for each program on OCG website.



Follow OCG [@NCIgenomics](#)

POSTER # 2224